

1. A group of researchers is running an experiment over the course of 30 months, with a single observation collected at the end of each month. Let  $X_1, \dots, X_{30}$  denote the observations for each month. From prior studies, the researchers know that

$$X_i \sim f_X(x),$$

but the mean  $\mu_X$  is unknown, and they wish to conduct the following test

$$H_0 : \mu_X = 0$$

$$H_a : \mu_X > 0.$$

At month  $k$ , they have accumulated data  $X_1, \dots, X_k$  and they have the  $t$ -statistic

$$T_k = \frac{\bar{X} - 0}{S_k/\sqrt{n}}.$$

The initial plan was to test the hypotheses after all data was collected (at the end of month 30), at level  $\alpha = 0.05$ . However, conducting the experiment is expensive, so the researchers want to “peek” at the data at the end of month 20 to see if they can stop it early. That is, the researchers propose to check whether  $t_{20}$  provides statistically discernible support for the alternative. If it does, they will stop the experiment early and report support for the researcher’s alternative hypothesis. If it does not, they will continue to month 30 and test whether  $t_{30}$  provides statistically discernible support for the alternative.

- (a) What values of  $t_{20}$  provide statistically discernible support for the alternative hypothesis?

```
alpha = 0.05
m20 = 20
discernible_boundary_lower = qt(alpha/2, df = m20-1)
discernible_boundary_upper = qt(1-alpha/2, df = m20-1)

discernible_boundary_lower
## [1] -2.093024

discernible_boundary_upper
## [1] 2.093024
```

If  $t_{20} > 2.093024$ , then there is statistically discernible support for the alternative and we would reject the null.

- (b) What values of  $t_{30}$  provide statistically discernible support for the alternative hypothesis?

```
alpha = 0.05
m30 = 30
boundary_lower = qt(alpha/2, df = m30-1)
boundary_upper = qt(1-alpha/2, df = m30-1)

boundary_lower
## [1] -2.04523

boundary_upper
## [1] 2.04523
```

If  $t_{30} > 2.04523$ , then there is statistically discernible support for the alternative and we would reject the null.

- (c) Suppose  $f_X(x)$  is a Laplace distribution with  $a = 0$  and  $b = 4.0$ . Conduct a simulation study to assess the Type I error rate of this approach.

**Note:** You can use the `rlaplace()` function from the `VGAM` package for R (Yee, 2010).

```
library(VGAM)
#??rlaplace
alpha = 0.05
m20 = 20
m30 = 30
t_20 = qt(1-alpha, df = m20-1)
t_30 = qt(1-alpha, df = m30-1)

simu_type1 = function(N = 10000, b = 4){
  false_positives = 0
  for (i in 1:N){
    x = rlaplace(30, location = 0, scale = b)
    x20 = x[1:20]
    t20 = (mean(x20) - 0) / (sd(x20)/sqrt(20)) # Centered at 0 so mu = 0

    if (t20 > t_20){
      false_positives = false_positives + 1
    }
    else {
      t30 = (mean(x) - 0) / (sd(x)/sqrt(30))
      if (t30 > t_30){
        false_positives = false_positives + 1
      }
    }
  }
  return(false_positives / N)
}

type1_rate = simu_type1()
type1_rate

## [1] 0.0773
```

The Type I error rate for this approach is 0.0735 or 7.35%

- (d) **Optional Challenge:** Can you find a value of  $\alpha < 0.05$  that yields a Type I error rate of 0.05?
2. Perform a simulation study to assess the robustness of the  $T$  test. Specifically, generate samples of size  $n = 15$  from the  $\text{Beta}(10,2)$ ,  $\text{Beta}(2,10)$ , and  $\text{Beta}(10,10)$  distributions and conduct the following hypothesis tests against the actual mean for each case (e.g.,  $\frac{10}{10+2}$ ,  $\frac{2}{10+2}$ , and  $\frac{10}{10+10}$ ).
- What proportion of the time do we make an error of Type I for a left-tailed test?
  - What proportion of the time do we make an error of Type I for a right-tailed test?
  - What proportion of the time do we make an error of Type I for a two-tailed test?
  - How does skewness of the underlying population distribution effect Type I error across the test types?

```
t_test_errors = function(N = 10000, func, true_mean, n = 15){
  left_error = 0
  right_error = 0
  two_error = 0
```

```

for (i in 1:N){
  x = func
  t_stat = (mean(x) - true_mean) / (sd(x) / sqrt(n))
  if (t_stat < qt(0.05, df = n - 1)){ # Checks the left-tailed test error rate
    left_error = left_error + 1
  }
  if (t_stat > qt(0.95, df = n - 1)){
    right_error = right_error + 1
  }
  if(abs(t_stat) > qt(0.975, df = n - 1)){
    two_error = two_error + 1
  }
}
return(c(Left-tailed = left_error/N, Right-tailed = right_error/N, Two-tailed = two_error/N))
}

beta_10_2 = t_test_errors(func = function rbeta(n, 10,2), true_mean = 10/12)

## Error in parse(text = input): <text>:20:24: unexpected '='
## 19:   }
## 20:   return(c(Left-tailed =
##                ^

```

## References

Yee, T. W. (2010). The VGAM package for categorical data analysis. *Journal of Statistical Software*, 32(10):1–34.